

WIESŁAW GLIŃSKI

Instytut Informacji Naukowej i Studiów Bibliologicznych, UW

## ONTOLOGIE. PRÓBA UPORZĄDKOWANIA TERMINOLOGICZNEGO CHAOSU

Próbowano przedstawić wybrane kwestie metodologiczne dotyczące ontologii, ich budowy i oceny w kontekście wyszukiwania informacji. Omówiono wybrane, formalne i nieformalne definicje ontologii.

### 1. WSTĘP

Pojęcie ontologii wiąże się często z filozofią i oznacza dział filozofii, w którym rozważa się własności przedmiotów związane ze „sposobem istnienia” (jak np. istnienie rzeczy, istnienie własności; istnienie realne intencjonalne itd.) Klasę przedmiotów o tym samym „sposobie istnienia” nazywa się często kategorią ontologiczną. Wyróżnia się zatem kategorie rzeczy oraz kategorie własności. Systemy filozoficzne różnią się między sobą liczbą i rodzajem wyróżnianych przez siebie kategorii ontologicznych. Zwraca się także uwagę na związek między logiką a ontologią, który polega na tym, że kategoriom syntaktycznym zmiennych przyjmowanym w teorii logicznej przyporządkowane są odpowiednie kategorie ontologiczne w modelu semantycznym tej teorii (*Mała Encyklopedia Logiki*, 1970, s.193). Przypomnijmy, że choć etymologia tego terminu sięga 1720 r (*Merriam-Webster's Collegiate Dictionary*, 1996), to jednak sama teoria jest znacznie starsza, bowiem wiąże się już z pracami Arystotelesa (IV p.n.e), a także Gottfrieda Leibniza, Immanuela Kanta, Bernarda Bolzano, Franza Brentano czy Kazimierza Twardowskiego lub Stanisława Leśniewskiego. Większość z nich traktuje jednak ontologię jako naukę stanowiącą o rodzajach i strukturach: obiektów, właściwości, zdarzeń, procesów, relacji i dziedzin rzeczywistości (Smith, 2004). My jednak będziemy traktowali ontologię jako dział inżynierii wiedzy. Wychodząc z założenia, że choć komunikacja między ludźmi jest jakimś punktem odniesienia, jesteśmy przekonani, że konieczne jest tworzenie autonomicznych systemów klasyfikacji i

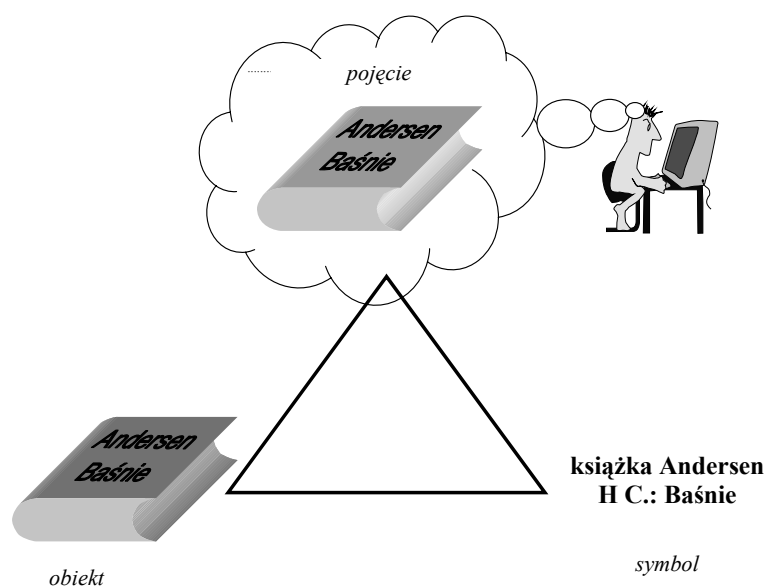
kategoryzacji pojęć będących swojego rodzaju metajęzykiem. Z punktu widzenia inżynierii wiedzy podkreśla się fakt, że ontologie powinny być z równą łatwością przetwarzane przez człowieka jak i przez maszynę. (Bassara, 2004). Dokonane w niniejszej pracy przemyślenia są w dużej mierze kontynuacją prac autora (por. Gliński, 2004).

Obecnie głównym problemem sieci WWW jest to, że wiele procesorów tekstu tworząc dokumenty zajmuje się wyłącznie zagadnieniami syntaktycznymi i to w sposób połowiczny. Metainformacje zawarte w stworzonych dokumentach niewiele mają wspólnego ze znaczeniem występujących tam informacji. Zdaniem Laurenta (1999) katastrofą podejścia typu WYSIWYG (*What you see is what you get*, *To co widzisz jest tym, co uzyskujesz*) było to, "że zwykły zbiór tekstowy stał się znacznie łatwiej przetwarzalny przez programy komputerowe niż dowolny dokument zaawansowanego edytora tekstowego czy programu do składu tekstu. W praktyce slogan WYSIWYG obecnie przekształcany jest na WYSIAYG: *What you see is all you get*. (*To co widzisz, jest wszystkim co możesz uzyskać*). Tekst jest do tego stopnia przeładowany informacjami dotyczącymi sposobu jego formatowania, że zabrakło miejsca na semantykę czy pragmatykę. (...) Formatowanie jest tym aspektem znaku, dzięki któremu wygląda on znacznie ładniej, ale jest podejściem absolutnie nie trafionym, jeśli chodzi o fundamentalne zagadnienie znaczenia tekstu." <sup>1</sup>

Andrzej Bassara (2004a) stwierdza, że pomiędzy wieloma czynnikami wpływającymi na jednoznaczność przekazu dwa zasługują na szczególną uwagę, nazwijmy je umownie: kategoryzacją oraz hierarchizacją. To pierwsze oznacza zdolność przyporządkowania symbolu, który pojawia się w komunikacie (np. "książka *Baśnie* Andersena") jakiejś ściśle określonej grupie obiektów, posiadających określone cechy (klasa wszystkich książek, pojęcie książki). Zestaw tych abstrakcyjnych grup, którymi posługuje się każdy z nas, a które zostały wykształcone w procesie edukacji możemy określić jako wewnętrzny model pojmowania świata. John F. Sowa (2000) wykorzystuje pojęcie trójkąta znaczeniowego (ang. *meaning triangle*) do zobrazowania tego procesu (**Rys. 1**). Choć komunikacja ludzka nie jest doskonała, to staramy się by ten abstrakcyjny model świata, jaki posiadamy stał się „sformalizowanym, samodzielny bytem” i by mógł stanowić punkt odniesienia dla odpowiednich stron przekazu, stając się odpowiednim metajęzykiem.

---

<sup>1</sup> Tekst przytoczony na podstawie artykułu J.F. Sowy (2000).

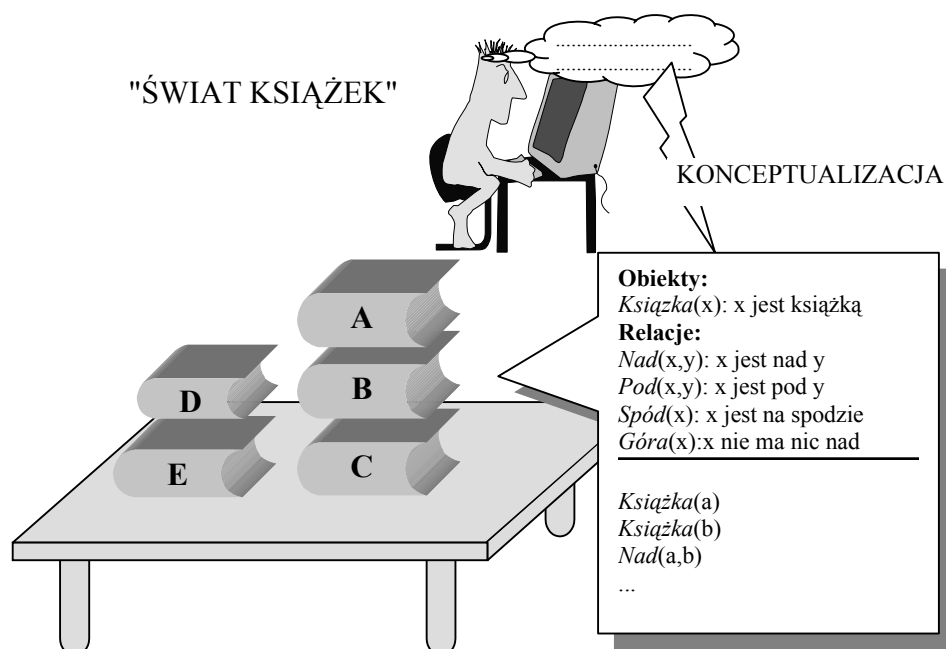


Rys. 1 "Trójkąt znaczeniowy" przedmiot-pojęcie – znak [na podstawie: Sowa, 2003]

## 2. DEFINICJE ONTOLOGII

Tom Gruber, który raczej dystansuje się od zaznaczonego we wstępie filozoficznego znaczenia terminu, definiuje ontologię jako wyraźną specyfikację konceptualizacji. Nie podaje jednak czym jest owa konceptualizacja. Aby lepiej zrozumieć kwestie ontologii jako związanej z inżynierią wiedzy pamiętajmy, że głównym jej celem jest dzielenie się wiedzą dziedzinową tak, aby była ona zrozumiała zarówno przez ludzi jak i maszyny.

Przyjrzyjmy się nieco bliżej kwestii konceptualizacji tak mocno zaznaczonej w większości opracowań. Posłużymy się w tym celu „światem książek” (**Rys. 2**) zainspirowanym przemyśleniami Nicola Guarina (1997).



Rys. 2 Konceptualizacja (ontologia) „Świat książek”.

Według autora możliwa konceptualizacja „świata książek” pokazanego na **Rys. 2** może przybrać postać zbioru uporządkowanego, składającego się z ze zbiorów obiektów oraz relacji (binarnych i unarnych):

$\langle \{a, b, c, d, e\}, \{ \} \rangle$

Wykorzystując logikę predykatów pierwszego rzędu spróbujemy dokonać wstępnego zapisu:

### Obiekty

Pięć obiektów tego samego typu; oznaczone są jako  $a, b, c, d, e$ .

$Książka(x)$  czytamy:  $x$  jest książką; podstawiając za zmienną  $x$  nazwy książek mamy:  $Książka(a)$ ,  $Książka(b)$ ,  $Książka(c)$ ,  $Książka(d)$ ,  $Książka(e)$ .

### Relacje

#### binarne:

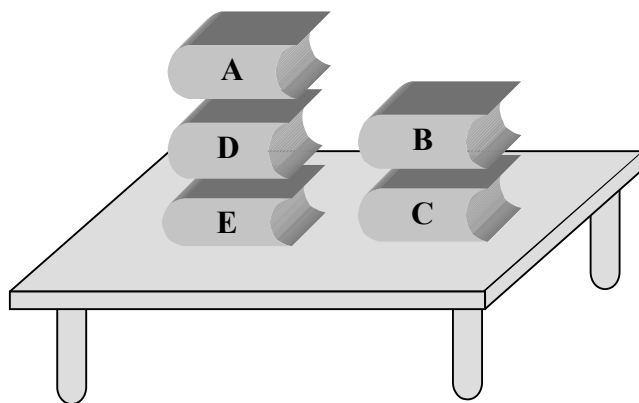
$Nad(x,y)$ :  $x$  jest nad  $y$

$Pod(x,y)$ :  $x$  jest pod  $y$

#### unarne:

$Spód(x)$ :  $x$  jest na spodzie

$Góra(x)$ :  $x$  jest na górze



Rys. 3 „Świat książek” – inny układ

Na **Rys. 3** widać inny układ książek. N. Guarino zadaje pytanie, „czy zmieniając układ książek zmienia się znaczenie terminów (Nad, Pod, Spód, Góra)”. Oczywiście, że nie. Zmienił się bowiem jedynie stan rzeczy. Znaczenie terminu "Nad" w obu przypadkach (**Rys. 2, Rys. 3**) jest takie samo, mimo że klocki te zmieniły swoje położenie. Cytując Guarino, możemy stwierdzić, że konceptualizacja jest zbiorem nieformalnych reguł, które w jakiś sposób ograniczają przedstawiany przez nas wycinek rzeczywistości, a które mogą być wykorzystane przez ludzi lub komputery w celu określenia odpowiednich obiektów i relacji; reguł, które mówią nam, czy dana książka jest nad innym czy nie, i to niezależnie od aktualnego stanu rzeczy.

N. Guarino (1995, 1997, 1998) badając literaturę przedmiotu w swoich pracach na temat ontologii podaje aż osiem definicji ontologii występujących w inżynierii wiedzy. Zdaniem Xaowei Yang (2004, s.1), definicje te jednak dotyczą trzech różnych aspektów ontologii: zawartości, formy i celu. Poniżej podajemy wybrane z nich:

- teoria na temat tego, jakie obiekty (encje) mogą istnieć w umyśle agenta;
- teoria dotycząca jakiejś dziedziny (a nawet konkretnych zadań, działań), opisująca pojęcia w sposób hierarchiczny (taksonomia) w celu ustalenie relacji semantycznych w danej dziedzinie;
- wyraźna specyfikacja konceptualizacji dla danej dziedziny na poziomie wiedzy, kierowana specyficznymi zadaniami;
- porozumienie na temat „wspólnej” konceptualizacji;
- logiczna teoria;
- teoria logiczna wprowadzająca ograniczenia do modeli logicznych;
- wyraźna, częściowa specyfikacja konceptualizacji, wyrażana z poziomu metajęzykowego, dotycząca zbioru odpowiednich dziedzin, której celem jest umożliwienie

projektowania modułowego, przeprojektowywania i ponownego wykorzystywania komponentów systemu wiedzy.

A. Bassara (2004a) powołuje się na najtrafniejszą jego zdaniem definicję ontologii z 2001 r., zaczerpniętą z pracy A. Maedche i S. Staaba (2001), określającą ontologię z formalnego punktu widzenia jako dwa zbiory: zbiór O definiujący strukturę ontologii oraz zbiór L definiujący słownik. Daje to jego zdaniem możliwość istnienia jednej ontologii, wspólnej dla różnych języków narodowych.

Zbiór  $O = \{C, R, Hc, rel, A\}$  określa strukturę pojęć, relacje między nimi, jak i teorię dotyczącą definiowanego modelu, gdzie:

C - stanowi zbiór wszystkich pojęć wykorzystanych w modelu. Pojęciem (często zwanym klasą) nazywamy ideę reprezentującą pewną grupę obiektów posiadających wspólną charakterystykę. Pojęcie w notacji identyfikowane jest przez symbol - najczęściej słowo;

R - jest zbiorem nietaksonomicznych relacji (w innych systemach zwanych właściwościami, atrybutami, ang. *slot*), definiowanych jako nazwane połączenie między pojęciami (np. jestCzęścią - oznacza, że jedno pojęcie występujące w relacji jest częścią drugiego). Należy w tym miejscu nadmienić, że relacja jest także pojęciem, niemniej jednak na potrzeby tej definicji oba zbiory winny być rozłączne;

Hc - stanowi zbiór taksonomicznych relacji pomiędzy konceptami;

rel - zdefiniowane nietaksonomiczne relacje pomiędzy pojęciami;

A - zbiór aksjomatów.

Zbiór  $L = \{Lc, Lr, F, G\}$ , który określa leksykon, czyli sposób, w jaki należy rozumieć pojęcia, w tym i relacje. Elementy tego zbioru to:

Lc - definicje leksykonu dla zbioru pojęć;

Lr - definicje leksykonu dla zbioru relacji;

F - referencje dla pojęć;

G - referencje dla relacji.

Literatura nt. sztucznej inteligencji zawiera wiele definicji ontologii, niektóre z nich stoją w opozycji do innych. My jednak wybieramy taką, która przyjmuje, że ontologia to formalny opis pojęć w danej dziedzinie wiedzy, tzw. zbiorze klas (*class*), czasem określanych jako pojęcia (ang. *concepts*)<sup>2</sup>, wraz z podaniem właściwości każdego z pojęć, czyli atrybutów (ang. *slot roles properties*). Ontologia wraz ze zbiorem indywidualnych wystąpień dla

---

<sup>2</sup> Dosłownie: formalna specyfikacja konceptualizacji.

każdego z pojęć składa się na bazę wiedzy. W rzeczywistości jednak istnieje cienka linia dzieląca ontologię od bazy wiedzy (por. np. Fernandez i in., 1997; Noy i McGuinness, 2004).

Przyjrzyjmy się kluczowym pojęciom. Klasy (ang. *class*) są pojęciem centralnym większości ontologii. Dzięki nim jesteśmy w stanie opisać pojęcia w danej dziedzinie wiedzy. Z kolei klasy mogą mieć podklasy, które odpowiadają pojęciom bardziej szczegółowym (podrzędnym) niż reprezentowane przez klasy. Tak więc możemy podzielić klasę wszystkich win na czerwone, białe i różowe lub dokonać innego podziału: na wina gazowane i niegazowane.

Atrybuty (ang. *slots*) opisują cechy klas i ich wystąpienia: Château Lafite Rothschild Pauillac ma cytrusowy bukiet smakowy i jest wyprodukowane przez winnicę Château Lafite Rothschild. Tak więc możemy stwierdzić, że mamy dwie własności win: „bukiet smakowy” z wartością "cytrusowy" oraz „producent” z wartością Château Lafite Rothschild. Zatem na poziomie klas możemy założyć, że wino może mieć własności opisujące jego zapach, smak, poziom cukru, producenta, winnicę itd.<sup>3</sup>

Ontologie wykorzystują teorie wywodzące się z algebry, teorii zbiorów, sieci semantycznych, oraz rachunków logicznych. Największy wpływ na istniejące języki tworzenia ontologii miały: rachunek predykatów (KIF, CycL), teoria ram (Ontolingua), oraz logika deskryptorów (CLASSIC) (Yang Xiaowei, 2004; Bassara, 2004a i 2004b; Maedche i Staab, 2001; Smith, 2004).

### 3. ZASADY TWORZENIA ONTOLOGII

Oczywiście tworzenie ontologii nie jest celem samym w sobie. Jest ono pokrewne definiowaniu zbioru danych i ich struktury w celu wykorzystania ich w innych programach. Systemy doradcze, programy niezwiązane z żadną konkretną dziedziną wiedzy oraz tzw. programowi agencji mogą wykorzystywać ontologie i bazy wiedzy zbudowane z ontologii jako zbioru danych. Na przykład, w cytowanym już „samouczku” F. Nathalyi Noy i Deborah L. McGuinness (2004) tworzona jest przykładowa ontologia potraw i win, dotycząca umiejętności właściwego połączenia wina z posiłkami.

Noy i McGuinness (2004, s.1) wymieniają też co najmniej pięć powodów tworzenia ontologii:

---

<sup>3</sup> Ontologia na podstawie przykładowej ontologii win dołączonej do programu PROTEGE 2000

- dzielenie się wspólnym rozumieniem ustrukturalizowanej informacji zarówno między ludźmi jak i maszynami (agentami internetowymi);
- umożliwienie wielokrotnego wykorzystania wiedzy z danej dziedziny;
- uczynienie założenia danej dziedziny wiedzy bardziej oczywistymi;
- oddzielenie dziedziny wiedzy od działań operacyjnych w danej dyscyplinie;
- analizowanie danej dziedziny wiedzy.

Jak słusznie podkreślili A. Fernandez, A. Gomez-Perez i N. Juristo (1997) oraz Yang Xiaowei (2004), tworzenie ontologii bardziej przypomina sztukę niż inżynierię. Budując systemy ontologii musimy pamiętać o jej cesze, jaką jest współdzielenie wiedzy. Stąd słuszne stają się zarzuty Basary (2004a), pod adresem ontologii rozumianej za Gruberem jako tylko „formalna specyfikacja konceptualizacji” (T. Gruber 1993, s.1). Według Bassary, „ontologia, zbudowana na potrzeby jednej aplikacji i wykorzystana tylko poprzez wewnętrzne procesy sterowania, nie ma prawa do nazywania się ontologią” (Basara, 2004a),.

Jednak już w 1993 T. Gruber zaproponował następujące kryteria projektowania systemów ontologii (Gruber, 1993, s. 2-3).:

- **j a s n o ś ć**: ontologia powinna w sposób efektywny przedstawiać zamierzone znaczenie definiowanych terminów. Definicje powinny być obiektywne, niezależne od kontekstu społecznego lub informatycznego. Wydaje się, że formalizm to umożliwia. Zatem, jeśli tylko to możliwe, definicje powinny być wyrażane przez aksjomaty logiczne. Tam, gdzie to możliwe, powinniśmy przekładać definicje pełne (predykaty definiowane przez warunki konieczne i wystarczające) nad definicjami cząstkowymi (predykaty definiowane przez warunki konieczne lub wystarczające). Oczywiście wszystkie definicje powinny posiadać opis w języku naturalnym;

- **s p ó j n o ś ć**: ontologia powinna być spójna, co oznacza, że powinien istnieć jakiś mechanizm inferencji. W końcu same aksjomaty powinny być logicznie spójne. Spójność powinna również stosować się do pojęć określanych w sposób nieformalny. Jeśli zdanie które można wyprowadzić z aksjomatów jest w sprzeczności z definicją lub podanym nieformalnie przykładem, to ontologia taka jest niespójna;

- **r o z s z e r z a l n o ś ć**: ontologia powinna być tak zaprojektowana, aby umożliwić wykorzystanie wspólnego słownika. Powinna zaoferować podstawy terminologiczne dla całego zakresu oczekiwanych zadań zaś reprezentacja powinna być tak przeprowadzona, aby możliwe było rozszerzanie i zawężanie ontologii w sposób monotoniczny. Innymi słowy,



powinna istnieć możliwość definiowania nowych terminów na podstawie istniejącego słownika w sposób, który nie wymaga rewizji istniejących już definicji;

- m i n i m a l n e z a a n g a ż o w a n i e s y m b o l i c z n e: konceptualizacja powinna być określona na poziomie wiedzy bez wykorzystywania jakiejś szczególnej symboliki. Minimalne zaangażowanie symboliczne oznacza, że wybór symboliki podyktowany jest wyłącznie względami ułatwiającymi proces notacji. Powody tego wynikają z faktu, że agenci systemowi mogą działać w zupełnie innym środowisku;

- m i n i m a l n e z a a n g a ż o w a n i e o n t o l o g i c z n e: ontologia powinna wymagać minimalnego zaangażowania ontologicznego, czyli powinna wprowadzać jak najmniej założeń i ograniczeń, a przy tym tylko takie, które są niezbędne dla systemów reprezentacji wiedzy, Jak bowiem podaje Xiaowei Yang (2004,s.10), im bogatszy słownik i im więcej jest ograniczeń, tym bardziej prawdopodobne jest, że w przyszłości nowe definicje mogą nie być zgodne z przyszłymi potrzebami reprezentacji.

A. Bassara (2004b) dokonując przeglądu literatury wymienia kolejne fazy tworzenia ontologii, w których odróżnia sam właściwy proces tworzenia ontologii od czynności go poprzedzających oraz takich procesów jak: wdrożenie, utrzymanie systemu oraz ocena ontologii:

- motywacja jako proces inicjujący;
- definicja zasięgu, czyli ustalenie jaki wycinek modelowanego świata będzie dotyczyła budowana ontologia;
- sam proces budowania ontologii, na który składa się:
  - identyfikacja pojęć,
  - budowanie struktury pojęć (na kształt drzewa, hierarchii),
  - modelowanie relacji jako atrybutów klas;
- ocena wyników;
- wdrożenie;
- utrzymanie.

My jednak postaramy się skupić na samym procesie budowania ontologii. F.N. Noy i D.L. McGuinness (2004) proponują następujące działania:

- definicje klas w ontologiach;
- ułożenie klas w hierarchiczne struktury (nadklasy – podklasy);
- zdefiniowanie własności klas i opisanie dopuszczalnych zbiorów wartości dla nich;

- wprowadzanie wartości dla poszczególnych własności dla odpowiednich wystąpień danych klas.

W swej pracy F.N. Noy i D.L. McGuinness (2004, s.4) podkreślają, że cały proces ma charakter powtarzalny, a ponadto zawsze istnieje możliwość wyboru jednego spośród wielu możliwych modeli. "Najlepszy model" to ten, który zależy od konkretnego zastosowania. W procesie wyboru odpowiednich pojęć należy dbać o to, aby były one możliwie najbliższe obiektom (w znaczeniu logicznym lub fizycznym), zaś relacje powinny odpowiadać danej dziedzinie. W konsekwencji obiekty winny być reprezentowane przez rzeczowniki a relacje przez czasowniki w zdaniach opisujących daną dziedzinę. Zgodnie ze zdrowym rozsądkiem, tworząc ontologię należy odpowiedzieć sobie na pytanie, do czego będziemy ją wykorzystywać oraz jak ogólna, względnie jak szczegółowa, ma być rozważana ontologia. F.N. Noy i D.L. McGuinness (2004, s.4) radzą, że w wyborze spośród wielu alternatywnych rozwiązań (ontologii) należy kierować się zasadą, że „lepszą” jest ta, która lepiej będzie współdziałała z naszymi aplikacjami i spełniała cele, które sobie obraliśmy ; będzie bardziej intuicyjna oraz wyczerpująca i łatwiejsza w utrzymaniu.

Po stworzeniu pierwszej wstępnej wersji ontologii możemy przystąpić do jej oceny przez zastosowanie jej w konkretnych aplikacjach doradczych w rozwiązywaniu problemów i/lub przedłożyć ją do oceny ekspertom dziedzinowym. Rezultatem takiej (takich) ocen (testowania) prawie zawsze jest konieczność zrewidowania ontologii. Należy dodać, że postępowanie takie jest procesem powtarzającym się, które towarzyszy nieustannie procesowi powstawania ontologii.

Prześledźmy siedem kroków w tworzeniu ontologii sugerowanych przez F.N. Noy i D.L. McGuinness (2004).

### **Krok 1: Ustalenie domeny oraz zasięgu ontologii**

Sugeruje się, aby rozwój ontologii rozpoczął się od zdefiniowania jej dziedziny oraz zasięgu. Innymi słowy należy odpowiedzieć na następujące podstawowe pytania:

- Jaka jest dziedzina (zbiór obiektów) tworzonej ontologii i jak szeroki zakres będzie brany pod uwagę w budowanej ontologii?
- W jakim celu zamierzamy użyć naszej ontologii?
- Na jakiego typu pytania będzie w przyszłości "odpowiadała" tworzona ontologia?
- Kto będzie używał i utrzymywał tworzoną ontologię?

Jednym ze sposobów ustalenia zasięgu budowanej ontologii jest podanie tzw. listy pytań wzorcowych, czyli pytań, na które winna odpowiedzieć baza wiedzy utworzona na podstawie ontologii. Pytania tego typu będą służyły jako swego rodzaju „papierek lakmusowy” w późniejszym etapie testowania ontologii. Aby zrozumieć naturę tego typu pytań należy odpowiedzieć sobie na pytania: "Czy ontologie zawierają wystarczająco dużo informacji, aby odpowiedzieć na pytania wzorcowe? Czy pytania te wymagają ustalenia jakiegoś specjalnego poziomu szczegółowości? Oczywiście pytania wzorcowe są tylko swego rodzaju „szkieletem” i nie muszą być w żadnej mierze wyczerpujące.

## **Krok 2: Wykorzystanie istniejących ontologii**

Autorzy publikacji na temat ontologii radzą, aby zastanowić się nad wykorzystaniem w całości lub części istniejącej systemów. Wykorzystanie istniejących ontologii może okazać się konieczne, jeśli przyszły system stosujący naszą ontologię będzie wchodził w interakcję z innymi systemami, które bazują na określonych systemach z kontrolowanym słownictwem. Wiele ontologii dostępnych jest w formie elektronicznej i może być bezpośrednio zaimplementowanych do systemu. Wykorzystywany formalizm, w którym wyrażona jest ontologia nie ma często większego znaczenia, gdyż wiele istniejących systemów posiada opcje importu i eksportu. Nawet jednak wtedy, gdy formalizm jednego systemu nie jest zgodny z innym, zadanie tłumaczenia nie jest zbyt trudne.

Przykładowo można wykorzystać bibliotekę Ontolingua dostępną pod adresem: <http://www.ksl.stanford.edu/software/ontolingua/>, ontologie DAML (<http://www.daml.org/ontologies/>), bogatą grupę publicznie dostępnych ontologii dla świata komercyjnego (np. UNSPSC (<http://www.unspsc.org/>), RosettaNet (<http://www.rosettanet.org/>), DMOZ ([www.dmoz.org](http://www.dmoz.org))) itd.

## **Krok 3: Ustalenie wszystkich najważniejszych terminów w projektowanej ontologii**

Pożyteczną czynnością jest niewątpliwie wymienienie wszystkich terminów, które będą wchodziły w skład budowanych przez nas zdań lub których znaczenie chcielibyśmy wytłumaczyć użytkownikowi. A zatem powinniśmy ustalić: Jakie to są terminy? Jakimi własnościami charakteryzują się obiekty reprezentowane przez nie? Co moglibyśmy powiedzieć o tych pojęciach?

Tak więc na wstępie ważną sprawą jest uzyskanie w miarę wyczerpującej listy terminów bez obawy, że niektóre z nich mogą się ze sobą pokrywać. Według Noy i McGuinness (2004) nie jest istotne, czy terminy te będą reprezentowały obiekty czy tylko relacje, czy też własności i czy odpowiednie pojęcia będą odpowiadały klasom czy tylko własnościom.

#### **Krok 4: Definiowanie klas i hierarchii klas**

W następnej kolejności powinniśmy zadbać o tworzenie (odzwierciedlenie) hierarchii obiektów oraz własności powiązanych ze sobą pojęć (ang. *slots*). Jak zauważyły Noy i McGuinness, nie jest łatwo stworzyć hierarchię bez ustalania własności obiektów. Najpierw stworzymy tylko kilka definicji terminów (pojęć) i układamy je w odpowiedniej hierarchii, a następnie przechodzimy do analizowania opisu własności tych pojęć itd. Podkreśla się, że są to jedne z kluczowych etapów w procesie tworzenia ontologii. Jak zauważyli M. Uschold i M. Gruninger (1996), istnieje kilka możliwych do wyboru strategii w budowie hierarchii.

Podejście „góra – dół” (ang. *top-down*) zaczyna się od definicji najbardziej ogólnych pojęć w danej dziedzinie, a w konsekwencji ich kolejnego uszczegółowiania. Zdaniem A. Bassary podejście to daje doskonałą kontrolę nad stopniem szczegółowości, jaki jest pożądany przez projektanta, jednak jego mankamentem może być włączenie do ontologii klas nadrzędnych, które niekoniecznie wymagane są przez końcowych użytkowników (Bassara, 2004b).

Kolejne podejście to „dół – góra” (ang. *bottom-up*). Jest to proces, który zaczyna się wraz z definicją najbardziej szczegółowych pojęć, obrazowo mówiąc „listków” z „drzewa hierarchii”. Po kolei przechodzimy poszczególne poziomy łącząc je w większe całości – bardziej ogólne pojęcia. Zdaniem Bassary wadą tego podejścia jest zbyt duża ilość detali oraz trudność w znajdowaniu klasy stanowiącej nadklasę dla dwóch znacząco różnych klas, co w rezultacie może doprowadzić do większej liczby poprawek (Bassara, 2004b).

Odminnym podejściem jest połączenie wymienionych wcześniej procesów: „góra - dół” oraz „dół - góra”. W takim przypadku staramy się zdefiniować najbardziej „rzucające się w oczy” (oczywiste) pojęcia, a następnie przechodzimy odpowiednio przez proces ich uszczegółowiania lub uogólniania.

Żadna z podanych metod nie jest lepsza od pozostałych. Wybór strategii zależy od osobistych preferencji tworzącego ontologie. Zdaniem wielu teoretyków podejście mieszane jest najłatwiejszą metodą. Którekolwiek podejście zostanie wybrane zawsze należy rozpocząć

od definiowania klas. Z listy terminów w kroku trzecim wybieramy terminy, które opisują niezależne istnienie obiektów. Terminy te będą klasami w ontologii i staną się punktami odniesienia w tworzonej hierarchii. Właściwe podejście zależy w dużej mierze od „osobistego” spojrzenia na daną dziedzinę. Jeśli projektant posiada wiedzę systematyczną na temat danej dziedziny, to zaleca się podejście „z góry na dół” (od ogółu do szczegółu). Najłatwiejszym podejściem według Noy i McGuinness (2004) może okazać się strategia łączona, gdyż pojęcia znajdujące się „po środku” hierarchii są przeważnie najbardziej opisowe. Niezależnie od wyboru strategii uogólniania, należy pamiętać o definiowaniu klas. Z listy przedstawionej w kroku trzecim, wybieramy terminy, które będą klasami w naszej ontologii i staną się niejako punktami zaczepienia tworzonego „drzewa hierarchii”.<sup>4</sup> Porządkujemy klasy w układ hierarchiczny, sprawdzając za każdym razem, czy dany obiekt (pojęcie), który jest wystąpieniem jednej klasy jest też w sposób konieczny (czyli z racji swej definicji) wystąpieniem jakiejś innej klasy. Jeśli klasa A jest nadklasą klasy B, to każde wystąpienie B jest także wystąpieniem A.

### **Krok 5: Definiowanie własności klas**

Same klasy nie dostarczają wystarczającej informacji, aby móc odpowiedzieć na pytanie wzorcowe z kroku pierwszego. W momencie, kiedy zdefiniowaliśmy niektóre z klas, musimy zdefiniować wewnętrzną strukturę pojęć.

Proces ustalania klas na podstawie listy terminów z kroku trzeciego mamy już za sobą. Tak więc większość pozostałych terminów wydaje się dobrymi kandydatami na własności klas.

Dla każdej grupy własności musimy ustalić, której klasy ona dotyczy. Cechy te stają się własnościami klas w ontologii.

Ogólnie istnieje wiele typów własności obiektów, wymienimy niektóre z nich:

- własności wewnętrzne (immanentne) “intrinsic”;
- własności zewnętrzne “extrinsic” ;
- własności typu „części” (ang. *parts*), kiedy dany obiekt składa się z części, co należy rozumieć zarówno w sensie fizycznym, jak i abstrakcyjnym;

---

<sup>4</sup> Noy and Deborah L. McGuinness (2004) proponują aby przedstawiać klasy jako predykaty unarne wtedy, gdy pytania formułowane pod adresem danego pojęcia posiadają tylko jedną zmienną, jak np. „Czy ten obiekt jest winem?” Zaś atrybuty jako predykaty binarne, które wyodrębniamy na podstawie pytań z dwoma niewiadomymi, jak np. w pytaniu: „Czy ten smak obiektu jest mocny?” lub „Jaki jest smak tego obiektu?”

- stosunki z innymi obiektami; dotyczy to relacji, jakie zachodzą między poszczególnymi obiektami;
- własności dziedziczone; wszystkie podklasy danej klasy dziedziczą jej własności, własność powinna być zatem dołączona do najbardziej ogólnej klasy.

#### **Krok 6: Definiowanie cech własności (atrybutów) klas**

Własności klas mają dodatkowo pewne cechy opisujące typ ich wartości: dozwolony ciąg wartości, liczbę wartości oraz inne dodatkowe cechy, które mogą wiązać się z własnościami. Tzw. kardynalność atrybutu określa, jak wiele wartości może mieć dana cecha. Niektóre systemy rozróżniają między własnością pojedynczą, która może mieć wyłącznie jedną wartość, a własnością wielokrotną, tj. przyjmującą więcej niż jedną wartość. Inne systemy pozwalają na określanie minimalnej oraz maksymalnej liczby dopuszczalnych wartości, co czyni ustalanie kardynalności znacznie bardziej precyzyjnym. Tzw. „fasetowość” (dopuszczalnych wartości dla atrybutów) określa jakiego rodzaju wartości może przybierać dana własność. W systemach, w których dołączamy własności do klas, klasy zwykle tworzą domenę danej własności. Nie ma więc potrzeby określania osobno domen. Zaleca się, aby „kandydatem” na domenę lub zakres danej klasy była klasa możliwie najbardziej ogólna.

#### **Krok 7: tworzenie wystąpień klas**

Ostatnim krokiem w konstruowaniu ontologii jest tworzenie wystąpień dla poszczególnych klas. Definiowanie indywidualnych wystąpień dla klas wymaga:

- wybrania klasy,
- stworzenia wystąpienia dla klasy,
- określenia własności.

### **4. ZAKOŃCZENIE**

Ontologie zyskały swoje miejsce w dziedzinie inżynierii wiedzy. Prowadzone są liczne badania na temat ich użyteczności i oceny. Powstały nawet standardy oceny ontologii. Wydaje się, że w kontekście tzw. Semantycznej Sieci WWW badania nad ontologią zyskują i będą zyskiwały coraz większą rolę. Sama konstrukcja ontologii jest jednak procesem niezwykle żmudnym, a przedstawione powyżej „wskazówki” dotyczące jej tworzenia (względnie pielęgnacji) dalekie są od doskonałości. Jednak badania nad tą dziedziną rozwijają

się w coraz większym stopniu, co niechybnie przyczyni się do znacznego jej usystematyzowania. Jak słusznie zauważył A. Bassara, tworzenie ontologii wraz ze wzrostem stopnia usystematyzowania przedmiotu przestaje być sztuką, a staje się w coraz większym stopniu inżynierią, bo tak samo jak inżynieria wymaga wyczucia i ogromu doświadczenia w realizacji tego typu przedsięwzięć (Bassara, 2004b).

Obecnie zwraca się uwagę na całe grupy zagadnień związanych z ontologiami: tworzenie konkretnych ontologii, analizę i ewaluację systemów ontologii, utrzymywanie ontologii, języki ontologii, narzędzia do tworzenia ontologii. Nadal jednak elementem kluczowym i sprawiającym najwięcej problemów jest pozyskiwanie wiedzy koniecznej w budowaniu ontologii. Wydaje się, że ponowne wykorzystywanie w całości lub części istniejących ontologii nieznacznie łagodzi ten problem. Powstaje jednak pokusa stworzenia automatycznych technik pozyskiwania wiedzy, które mogłyby m.in.:

- wykorzystywać narzędzia lingwistyczne pozyskujące wiedzę na podstawie analizy tekstu;
- wykorzystywać proces uczenia się maszyn np. tworząc ontologie na podstawie ustrukturalizowanych dokumentów XML;
- badać strukturę sieci WWW, tworzenie ontologii na podstawie przeglądania ustrukturalizowanych zasobów WWW;
- wprowadzać swoistego rodzaju szablony do pozyskiwania wiedzy, w przypadku których eksperci określiliby tylko część koniecznej wiedzy.

W przypadku narzędzi i języków ontologii zwraca się uwagę na właściwy poziom ekspresyjności (wyrazistości) języka, poziom jego semantyki i liczbę założeń, jakie są przyjmowane. Powstaje też problem oceniania ontologii. W założeniu ontologia ma charakter subiektywny, w jakiś sposób odwzorowuje czyjąś wiedzę i rozumienie pojęć. Dlatego wydaje się jednak, że najlepszym sposobem na ocenę danej ontologii jest umieszczenie jej w konkretnej aplikacji i przetestowanie funkcjonowania całości.

## LITERATURA

Bassara, A (2004a). I weź tu dogadaj się – Ontologie [online]. *Gazeta IT* nr 1(20) [dostęp: 30 04. 2005]

Dostępny w WWW: [http://www.gazeta-it.pl/zw/git20/i\\_wez\\_tu\\_dogadaj\\_sie\\_ontologie.html](http://www.gazeta-it.pl/zw/git20/i_wez_tu_dogadaj_sie_ontologie.html)

Bassara, A. (2004b). Ontology Engineering – Ontologie [online]. *Gazeta IT* nr 2(21) [dostęp: 30 04. 2005]

Dostępny w WWW: [http://www.gazeta-it.pl/zw/git21/inzynieria\\_ontologii.html](http://www.gazeta-it.pl/zw/git21/inzynieria_ontologii.html)

- Fernandez, M.; Gomez-Perez, A.; Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In: *The AAAI-97 Spring Symposium Series on Ontological Engineering*, p. 33-40, March 1997.
- Gliński, W. (2004). Kwestie metodyczne projektowania ontologii w systemach informacyjnych. W: *Strategie informatyzacji i zarządzanie wiedzą*. Red. Zdzisław Szyjewski, Jerzy S. Nowak, Janusz K. Grabara. Warszawa: Wydawnictwa Naukowo-Techniczne, s.201-212
- Gruber, T.R. (1993). *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. Technical Report KSL 93-04 [online]. Knowledge Systems Laboratory, Stanford University. [dostęp: 10.05.2005] Dostępny w WWW: <http://tomgruber.org/writing/onto-design.pdf>
- Guarino, N. (1998). Formal Ontology and Information Systems. [online] Laboratory for Applied Ontology [dostęp: 10 05. 2005] Dostępny w WWW:<http://www.loa-cnr.it/Papers/FOIS98.pdf>
- Guarino, N. (1997). Understanding, Building, and Using Ontologies. *International Journal of Human Computer Studies*, vol. 46 no. 2/3, s. 293 - 310.
- Guarino, N.; Carrara, M.; Giaretta, P. (1995). Ontologies and knowledge bases: towards a terminological clarification. In: N. Mars (ed.). *Towards Very Large Knowledge Bases, Knowledge Building and Knowledge Sharing*. Amsterdam: IOS Press, p. 25-32.
- Laurent St., Simon (1999). *XML: A Primer, second edition*. Foster City, CA : M & T Books,
- López Fernández, M. (1999). Overview Of Methodologies For Building Ontologies [online] Departament de Llenguatges i Sistemes Informatics, Universitat Politècnica de Catalunya [dostęp:10 05. 2005]Dostępny w WWW:<http://www.lsi.upc.es/~bejar/aia/aia-web/4-fernandez.pdf>
- Maedche, A.; Staab, S (2001). Ontology learning for the Semantic Web. *IEEE Intelligent Systems* vol. 16 no2
- Mała Encyklopedia Logiki* (1970). Wrocław, Warszawa, Kraków: Zakład Narodowy Ossolińskich
- Merriam-Webster's Collegiate Dictionary* (1996). Merriam-Webster's [CD-ROM]
- Nahotko, M. (2003). *Semantyczny Web i jego ontologie* [online] *Biuletyn EBIB* nr 9 (49) [dostęp: 10 05. 2005] Dostępny w WWW: <http://ebib.oss.wroc.pl/2003/49/nahotko.php>
- Noy, F. N.; McGuinness, D.L. (2004). *Ontology Development 101: A Guide to Creating Your First Ontology* [online] Knowledge Systems Laboratory, Stanford University.. [dostęp:10 05. 2005] Dostępny w WWW: <http://www.ksl.stanford.edu/people/dlm/papers/ontologyDevelopment101.pdf>
- Smith, B. (2004). *Ontology and Information Systems* [online] State University of New York at Buffalo, Department of Philosophy, Ontology. [dostęp: 13.04.2004]. Dostępny w WWW: [http://ontology.buffalo.edu/ontology\(PIC\).pdf](http://ontology.buffalo.edu/ontology(PIC).pdf)
- Smith, B. (2004). Ontology and Information Systems [online]. *Stanford Encyclopedia of Philosophy* [dostęp: 10 05. 2005] Dostępny w WWW: <http://plato.stanford.edu/>
- Sowa, J. F. (2000). *Ontology, Metadata, and Semiotics* [online]. John F. Sowa and Cora Angier Sowa Website [dostęp: 10 05. 2005] Dostępny w WWW: <http://www.bestweb.net/~sowa/peirce/ontometa.htm>
- Uschold, M.; Gruninger, M. (1996). Ontologies: Principles, Methods and Applications. [online]Ontology -Based Knowledge Management [dostęp: 10 05. 2005] Dostępny w WWW: <http://bingo.crema.unimi.it/ontology/doc/ontology/uschold96ontologie.pdf>
- Yang Xiaowei (2004). *Ontologies and How to Build Them* [online]. [Advanced Network Architecture Group MIT Laboratory for Computer



Science, Xiaowei Yang home page] [dostęp: 13.04. 2004] Dostępny w WWW:  
<http://cordelia.lcs.mit.edu/~yxw/publications/area-exam.ps>

## ABSTRACT

We attempt to present selected methodical issues concerning designing and evaluating ontology in the context of information system retrieval. There were presented several formal and informal definitions of ontology.